

L. Leurs, 2001

(traducción: Gustavo Sánchez Muñoz)

## **CONTENIDO**

- 1) La historia del PDF
- 2) Las fuentes en los archivos PDF
- 3) La compresión en el formato PDF
- 4) El texto en los archivos PDF
- 5) PDF frente a otros formatos de archivo.
- 6) Microsoft Word y el formato PDF



# La historia del PDF

[L. Leurs](#), 2001

(traducción: Gustavo Sánchez Muñoz)

"La oficina sin papeles". ¿Recordais aquella pseudo profecía que nunca se desvanecía del todo, incluso aunque la historia se empeñara en demostrar que el uso de ordenadores ha aumentado el uso del papel?

PDF comenzó con el sueño de una oficina sin papeles. Era el proyecto mimado de uno de los fundadores de Adobe, John Warnock. Al comienzo hubo en Adobe un proyecto interno de crear un formato de ficheros para que los documentos se pudieran distribuir por la compañía, viéndose en cualquier ordenador, fuera cual fuera su sistema operativo.

Adobe ya disponía de dos cosas que casi encajaban con la idea: El PostScript (una tecnología independiente del dispositivo y la plataforma que servía para describir documentos) y Adobe Illustrator (un programa que funcionaba en distintas plataformas —bueno, sólo en Macintosh y Windows, el 99,9% de los ordenadores— y que podía abrir y mostrar ficheros PostScript muy simples) Los ingenieros de Adobe mejoraron ambas tecnologías y crearon un nuevo formato de ficheros (Portable File Document: PDF, una especie de PostScript optimizado) y una serie de aplicaciones capaces de crear y mostrar esos ficheros.

## PDF 1.0

la primera vez que Adobe habló de esa tecnología fue en la conferencia de Seybold, en San José (California), en 1991. En ese momento, se la mencionó como "IPS", siglas de Interchange PostScript ("PostScript de Intercambio"). La versión 1.0 de PDF se anunció en Comdex en otoño de 1992, donde esta tecnología ganó un Premio Comdex. Las

herramientas para crear y mostrar ficheros PDF se hicieron públicas a mediados de 1993. La primera versión carecía de interés para el mundo de la preimpresión. Ya disponía de enlaces internos y marcadores. Las fuentes se podían incrustar, pero el único espacio de color que admitía era RGB.

El nombre en código original para lo que posteriormente se convertiría en Acrobat fue "Carrusel" (*Carousel*). Por eso, los primeros ficheros PDF en Macintosh eran ficheros "CARO".

Adobe puso un precio astronómico a las herramientas para crear PDFs e incluso había que pagar 50\$ USA por el lector de Acrobat (*Acrobat reader*). Este planteamiento no ayudó precisamente a que PDF se convirtiera en algo muy popular de la noche al día. Posteriormente, Adobe redujo mucho el precio de Acrobat y lanzó una versión gratuita del lector.

## PDF 1.1

Acrobat 2 se hizo público en septiembre de 1994. Admitía la nueva definición 1.1 del formato PDF, que añadía:

- Vínculos externos.
- Enlaces encadenados.
- Características más seguras.
- Color independiente del dispositivo.
- Anotaciones.

En sí mismo, Acrobat adquirió algunas mejoras muy notables, que incluyó una nueva arquitectura para Acrobat Exchange que le permitía incorporar extensiones (*plug-ins*) y la capacidad de buscar dentro de ficheros PDF.

La propia firma Adobe fue uno de los primeros grandes usuarios de PDF. Pasaron a distribuir todos los documentos destinados a los desarrolladores en formato PDF. Otro de los primeros adeptos del formato PDF fue el departamento fiscal de los Estados Unidos, que comenzó a distribuir formularios en formato PDF.

Acrobat 2.1 incorporó capacidad multimedia y la posibilidad de añadir datos de audio y vídeo a los ficheros PDF.

ARCHIVOS Y PDF.

por aquellas fechas, PDF no era el único intento de crear un formato de ficheros portátiles independiente de los sistemas operativos y de los dispositivos. Su principal competidor era un producto llamado [Common Ground](#).

En 1995, Adobe comenzó a distribuir Acrobat Capture por el exorbitante precio de 4.000\$ US. Al mismo tiempo, Adobe comenzó a añadir capacidad de tratar PDFs en muchas de sus propios programas, incluidos frameMaker 5.0 y PageMaker 6.

## PDF 1.2, el mundo de la preimpresión se despierta

En 1996, Adobe lanzó Acrobat 3.0 (nombre en código "Ambar" —*Amber*—) y la correspondiente especificación PDF 1.2. Ésta fue la primera versión de PDF que verdaderamente tenía alguna utilidad para la preimpresión. Además de formularios, se incluyeron las siguientes facetas relacionadas con la preimpresión:

- Soporte de las especificaciones OPI 1.3.
- Soporte del espacio de color CMYK.
- Los colores directos ya se podían conservar dentro de un PDF.
- Las funciones de trama de semitonos y los valores de sobreimpresión ya se podían incluir.

la inclusión de un plugin para poder ver ficheros PDF dentro del navegador Netscape Navigator aumentó la popularidad de los ficheros PDF en una Internet que estaba en plena eclosión. Además, Adobe incluyó la posibilidad de enlazar ficheros PDF con documentos HTML y viceversa. El mercado de la impresión digital en blanco y negro comenzó a usar el formato PDF como base para imprimir en máquinas xerocopiadoras.

Con Acrobat 3, la arquitectura abierta de Acrobat Exchange comenzó a pagar dividendos y hacia 1997 y 1998 comenzaron a aparecer una serie de extensiones para preimpresión extremadamente interesantes, entre ellas varias herramientas esenciales para ese mercado. Entre ellas, destacaron

PrintShop y CheckUp (de [EMUCUS](#)), y CrackerJack (de [Lantana](#)). Agfa fue la primera compañía de peso que promovió el uso de PDFs para la impresión comercial a todo color con el lanzamiento de su sistema Apogee en 1998. Otros fabricantes se animaron poco después.

Aunque las firmas involucradas se esforzaron por dar impulso a PDF, el mercado reaccionó de forma un poco remisa. Esto se debió sobre todo al hecho de que el uso de PDF requiriera herramientas adicionales y cierto conocimiento del formato PDF, sus limitaciones y peculiaridades. La gente se quedaba un poco frustrada cuando descubría que PDF era un estándar muy abierto. Esto quería decir que, aunque PDF era utilizable en entornos de preimpresión, había simplemente demasiadas maneras en las que se podía crear un fichero PDF perfectamente válido e *inutilizable*.

## PDF/X-1: Un estándar que se impone (muy) lentamente

Para solucionar ese problema, un consorcio de compañías de preimpresión se reunieron en 1998 y acordaron la especificación PDF/X-1. Ésta está basada en las especificaciones de PDF 1.2 pero con la particularidad de que forman una descripción muy bien definida de cómo debe ser un fichero PDF para permitir transferencias sin problemas. Un fichero PDF/X-1 es un fichero del que puedes estar seguro que se han incluido todas las fuentes, en el que todas las imágenes de alta resolución se han incrustado, etc...;

Aunque PDF/X-1 se basa en PDF 1.2, se le añadieron una serie de operadores extra. Están descritos en la Nota Técnica de Adobe nº 5.188 e incluyen:

- la posibilidad de incrustar datos extra como es el caso de ficheros Copydot.
- Soporte para colores basados en ICC.
- La definición de sangre, corte y caja.
- Una clave que indica si el documento ya ha recibido reventados (*trapping*).

[Aquí](#) Puedes encontrar más información sobre PDF/X-1.

## PDF 1.3: Escuchar las necesidades de preimpresión

Adobe Acrobat 4, lanzado en abril de 1999 trajo consigo PDF 1.3. Estas nuevas especificaciones para PDF incluían:

- Fuentes CID de 2 bytes.
- Especificaciones OPI 2.0.
- Un nuevo espacio de color llamado DeviceN para mejorar el soporte de colores directos (*spot colours*)
- "smooth shading", una tecnología que permite degradados (*blends*, transiciones de color) muy suaves y eficientes.
- Anotaciones.

Adobe Acrobat también tenía novedades como programa, entre ellas:

- Soporte para páginas entre 5.080 × 5.080 mm y 1.143 × 1.43 mm.
- Webcapture [grabación en PDF parcial o total de sitios "web"].
- Configuraciones prefabricadas de parámetros de trabajo para Acrobat Distiller (lo que hacía mucho más fácil la creación de ficheros PDF correctos).
- Un cambio bastante confuso de nombres: Acrobat Exchange pasó a llamarse "Acrobat", lo que causa problemas al ser también el nombre del conjunto de programas.
- Facilidad de integración con Microsoft Office.

La versión inicial de Adobe Acrobat 4 (la llamada 4.0) contenía un buen número de errores de programación (*bugs*) que limitaban la utilidad del programa para usos de preimpresión. Muchos usuarios se irritaron bastante cuando Adobe intentó cobrar por la solución a esos errores con una actualización a 4.05. Por suerte, Adobe prestó atención a sus usuarios y envió una copia a los usuarios registrados (en Europa tuvimos que esperar unos cuatro meses aproximadamente).

Aunque aún había cuestiones relacionadas con el soporte en PDF para preimpresión que precisaban arreglo, es difícil no reconocer que PDF se había convertido en un estándar del mercado. En la Red se

descargaron más de 100 millones de copias de Acrobat Reader.

## Illustrator 9 y PDF 1.4: Acrobat tendrá que esperar

A mediados de 2000, Adobe hizo algo extraño: Lanzó Illustrator 9. Si bien sacar al mercado una nueva versión de un programa de ilustración no es nada raro, Adobe le añadió una faceta curiosa. Era la primera aplicación que soportaba PDF 1.4 y sus posibilidades de transparencias. Era la primera vez que Adobe no lanzaba conjuntamente una nueva especificación de PDF y una nueva versión de Acrobat. Además, Adobe no hizo completamente públicas las especificaciones de PDF 1.4, aunque en el documento técnico de Adobe nº 5.407 se documentaba el soporte de transparencias en PDF 1.4.

## Acrobat 5 (nombre clave: 'Brasil')

En mayo de 2001, Acrobat 5 salió por fin a la luz y PDF 1.4 se convirtió en una realidad. Lo cierto es que el formato PDF no había cambiado tanto. Con respecto a la preimpresión, los aspectos más relevantes eran:

- La inclusión de soporte para transparencias que permiten que un objeto (texto o imágenes) deje ver parcialmente lo que hay debajo.
- Una mejora en la seguridad, incluyendo encriptamiento de 128 bits y la opción de poder bloquear la calidad de impresión (se puede definir que un PDF sea sólo imprimible como baja resolución, por ejemplo).

Para usuarios no relacionados con la preimpresión, Adobe añadió algunas mejoras a PDF 1.4:

- El formato amplió la admisión de JavaScript (incluido JavaScript 1.5) y la integración con bases de datos.
- Los ficheros PDF "marcados" (*tagged*), que son ficheros PDF que contienen información estructural sobre los datos representados en un fichero PDF. Esto quiere decir que meta-

mentos PDF. Esto quiere decir que meta-información del tipo de títulos, bloques de texto, etc...; puede formar ahora parte del documento PDF.

Esto hizo mucho más fácil crear ficheros PDF que se podían adaptar por sí mismos al dispositivo en el que se estuviera usando. Esta característica estaba enfocada sobre todo al emergente mercado de publicaciones electrónicas (*e-books*), ya que permitían que los ficheros PDF se usaran en múltiples sistemas. Adobe comenzó a distribuir una versión de Acrobat Reader que funcionaba ya en los PDAs con PalmOS.

En conjunto, esto hacía más fácil reformar el contenido para distintas distribuciones y soportes.

La mayoría de los usuarios quedaron más complacidos con las nuevas características de Adobe Acrobat. Los usuarios de preimpresión contaban con las siguientes mejoras:

- Acrobat ya podía mostrar sobreimpresiones por sí mismo. De este modo, si un usuario colocaba texto amarillo sobreimpreso en un fondo cian, Acrobat 5 lo mostraba como un texto verde allí donde pisara. Hay que destacar que esta opción venía desactivada por omisión.
- Con Acrobat 5 ya se podían hacer operaciones por tandas (*batches*), lo que permitía, entre otras cosas, exportar carpetas enteras de ficheros PDF al formato EPS.
- Distiller ya podía comprimir imágenes que usaran colores definidos como DeviceN. Esto hizo que los ficheros PDF que contuvieran colores directos fueran mucho más reducidos.
- Acrobat y Distiller mejoraron su motor de gestión del color, conocido como "ACE" (Adobe Color Engine), lo que proporcionaba un control mucho más preciso.
- Las anotaciones en ficheros PDF eran más flexibles y podían hacerse ya a través de Internet.

Acrobat 5 fue una actualización de mucho valor para los usuarios no relacionados con la preimpresión:

- Las características disponibles para formularios mejoraron muchísimo.
- El interfaz de uso de Adobe Acrobat recordaba



La historia de uso de Adobe Acrobat incorporaba ahora más a las aplicaciones de Microsoft Office. La integración de Adobe Acrobat con esta *suite* mejoró notablemente.

- Ya se podía usar Distiller o Catalog desde dentro de Acrobat.
- Cuando se abría un fichero en Acrobat, las miniaturas se creaban ya automáticamente.
- Adobe Acrobat dispuso de más y mejores filtros de exportación, incluida una opción para exportar datos como formato de texto RTF. Esto facilitó mantener la apariencia de los ficheros al exportarlos a procesadores de texto.
- Curiosamente, para algunos usuarios, Acrobat 5 fue un paso atrás, ya que Adobe reemplazó la extensión "Paper Capture" (que podía efectuar reconocimiento óptico de caracteres (OCR) de páginas escaneadas para crear PDF basados en verdadero texto) por un servicio en línea muy limitado.

---

## PDF 1.5 y Acrobat 6: Aun más elecciones para los ya confusos usuarios

En abril de 2003, Adobe anunció el lanzamiento de Acrobat 6, que salió a la venta en mayo de ese mismo año. El nombre clave de Acrobat 6 fue 'Newport'. Como es costumbre, la aparición de una nueva versión de Adobe Acrobat conllevó la de un nuevo nivel en el formato PDF, la versión 1.5.

En el nivel PDF 1.5 se han añadido nuevas características que los distintos programas tardarán probablemente bastante tiempo en poder aplicar o siquiera reconocer. Entre las nuevas cosas se incluye:

- Mejora los sistemas de compresión y admite algunos nuevos, como el sistema JPEG 2.000.
- Admite y reconoce capas (*layers*).
- Mejora el trabajo con ficheros PDF 'marcados' (*tagged PDF*).

El programa Adobe Acrobat en sí ofrece algunas cosas aparte del nuevo nivel de PDF.

El lector pasa a llamarse sólo 'Adobe Reader' e

incluye las funciones del antiguo 'Adobe eBook Reader'. Desgraciadamente eso ha hecho que el tamaño del programa haya crecido y que un fichero sea ahora hasta 1.000 veces mayor que cualquiera de los típicos documentos PDF de trabajo en oficinas.

Adobe Acrobat 6 Profesional es la versión de alto nivel destinada a usuarios del mundo de la preimpresión, y cuenta con un gran abanico de nuevas características.

- Capacidad de chequeo previo (*preflight*).
- Optimización de ficheros PDF.
- Reglas y guías.
- [Job tickets](#).
- Trabajo con el estándar [PDF/X](#)
- Previsualización de separaciones de color.
- Acoplador de transparencia.
- Capas
- Herramientas de medida (regla) y lupa.
- Un nuevo interfaz de usuario similar al de los otros programas de Adobe.

INICIO EL RINCÓN DEL PDF





# Las fuentes en los ficheros PDF

[L. Leurs](#), 2000  
(traducción: Gustavo Sánchez Muñoz)

En el formato PDF se pueden usar fuentes PostScript (tipo 1 y tipo 3) y fuentes True Type. Estas fuentes se incluyen preferentemente en los ficheros PDF para que se puedan ver e imprimir tal cómo fueron hechos por sus creadores. Si faltan algunas fuentes, Acrobat (ya sea su versión completa o Acrobat Reader) intentará emularlas usando alguna de las fuentes Multiple Master que llevan incluidas. De este modo, el documento no se representará tan exactamente como su creador quiso, pero, al menos, el texto no refluirá.

Las fuentes Multiple Master (MM) que Adobe Acrobat usa son:

- Adobe Serif MM.
- Adobe Sans MM.

## Qué fuentes se usan en un PDF

Acrobat (ya sea su versión completa o Acrobat Reader) ofrecen una opción de información sobre las fuentes es su menú "Archivo - Datos del documento - Fuentes". En este apartado se listan sólo aquellas fuentes usadas en la página activa en ese momento. No debes usarlo como una fuente de información sobre todas las fuentes usadas en un fichero PDF. Para obtener una lista completa de todas las fuentes usadas en un fichero PDF debes usar una extensión del tipo de PitStop (de [Enfocus](#)).

## Incrustar fuentes

Acrobat Distiller ofrece la opción de añadir

automáticamente las fuentes perdidas a los ficheros PostScript que tiene que procesar. En el menú de "Configuración - Ubicación de fuentes", el usuario puede definir varias carpetas para que Acrobat Distiller busque y sepa dónde localizar las fuentes que falten en el fichero PostScript. Para los usuarios de Macintosh: EL programa sólo necesita las fuentes de impresora, no las de pantalla.

## Qué fuentes no se incluyen en los ficheros PDF

Hay algunas diferencias fundamentales entre Acrobat 3.x y Acrobat 4.x en la forma de gestionar las fuentes.

Hay 14 fuentes que Acrobat Distiller 3.x nunca incrusta en un fichero PDF. Son:

- Courier, Courier-Bold, Courier-Oblique, Courier-BoldOblique.
- Times-Roman , Times-Bold , Times-Italic, Times-BoldItalic.
- Helvetica, Helvetica-Bold, Helvetica-Oblique, Helvetica-BoldOblique.
- Symbol.
- ZapfDingbats.

Estas fuentes (ZapfDingbats excluida) son las llamadas 13 fuentes básicas.

Adobe Acrobat 4.x no tiene problemas en incrustar esas 14 fuentes. Pero sí tiene otra restricción: Si las normas de licencia (*licensing policy*) de una fuente True Type impiden que la fuente de incruste en un fichero, Adobe Acrobat 4.x respetará esta restricción y no incrustará la fuente. [Lee esto](#) si quieres saber más sobre este problema.

## Incrustar subjuegos de fuentes

También es posible incluir sólo aquellos caracteres de una fuente que se usen en una publicación. Esta técnica se llama "incrustar subjuegos de fuentes". Al usar la opción "Configuración - Opciones de trabajo - Fuentes - Incrustar subjuegos de fuentes por debajo de XXX %" de Acrobat Distiller, el usuario puede decidir cuántos caracteres deben haberse usado en una publicación para que se incruste una porción de esa fuente.

para que se incluyese una porción de esa fuente.

incrustar subjuegos tiene dos ventajas:

- El tamaño del fichero PDF final se reduce y esto puede ser muy útil si el tamaño de un PDF es verdaderamente crucial, como es el caso de PDFs que deban situarse en la Red.
- Un RIP usará siempre una fuente incrustada como subjuego incluso aunque disponga por sí mismo de la fuente completa. Esta es una manera de evitar recolocaciones extrañas de textos debido a diferencias menores entre versiones de una misma fuente.

Sin embargo, incrustar subjuegos también tiene dos desventajas:

- Si el usuario opta por editar algo de texto en Adobe Acrobat y un carácter que se necesite no ha sido incrustado en el subjuego, no podrá usarse.
- Fusionar dos ficheros PDF que contienen subjuegos diferentes de una misma fuente puede hacer que desaparezcan caracteres necesarios.

[INICIO](#) [EL RINCÓN DEL PDF](#)





# La compresión en el formato PDF

[L. Leurs](#), 2001

(traducción: Gustavo Sánchez Muñoz)

Los ficheros PDF son mucho menores que sus equivalentes PostScript. Aunque esto se consigue en parte gracias a una mejor estructura de los ficheros, la razón principal se debe a los algoritmos de compresión extremadamente eficaces que admiten los PDFs.

Todos los textos y todas las órdenes PostScript que forman parte de un PDF se pueden comprimir usando el algoritmo LZW (Lempel-Ziv-Welch). Esta compresión por sí sola ya puede dejar el tamaño de un fichero PDF en la mitad del tamaño de su equivalente PostScript.

Si la compresión está activada en el Acrobat Distiller, lo primero que hará este programa es descomprimir todas las imágenes que se hallen dentro del fichero PostScript y luego las volverá a comprimir al crear el fichero PDF. En el caso de las imágenes, existen varios algoritmos de compresión admisibles: JPEG, Zip, CCITT, RLE y "hinchar/deshinchar" (*flate*).

Para comprobar qué tipo de compresión se ha utilizado, basta con abrir el fichero PDF en un simple editor de texto que admita datos binarios (el Bloc de Notas, etc...) y buscar allí la cadena `"/Filter"`.

## La compresión JPEG

La compresión JPEG se usa para imágenes de cuatricromía y escala de grises. Es un algoritmo de compresión que puede ser "con pérdidas" (*lossy*) y "sin pérdidas" (*non-lossy*). Sin embargo, en Adobe Acrobat sólo está disponible el modelo "con pérdidas". Esto quiere decir que algunos detalles de

la imagen se pierden o deterioran cada vez que se la comprime. Cuanto mayor es la tasa de compresión, mayor es el deterioro de las imágenes. Si quieres saber más sobre el algoritmo de compresión JPEG, [puedes leer esto](#) [y también [esto otro](#), en inglés].

Acrobat Distiller 3.x ofrecía cinco niveles de compresión JPEG:

1. Baja: Proporciona una pérdida de calidad que es generalmente aceptable salvo en el caso de los trabajos más exigentes. La tasa media de compresión es de 1:2.
2. Baja - media
3. Media: Proporciona una tasa de calidad para trabajos de tipo medio. La tasa media de compresión es de 1:5.
4. Alta - media
5. Alta: not acceptable for prepress any more.  
Average compression ratio: 1/10

Acrobat Distiller 4 tenía disponibles ajustes muy similares, pero no se correspondían exactamente con los de Acrobat 3. En Acrobat Distiller 4.0 había además un fallo de programación que afectaba al algoritmo JPEG. Este fallo se arregló con la actualización a 4.05. Las opciones son:

1. Mínima.
2. Baja.
3. Media.
4. Alta.
5. Máxima.

Las imágenes que se han comprimido usando el algoritmo JPEG no deben recomprimirse, ya que esto causará una pérdida acumulativa de calidad. Si se destila un fichero que contiene imágenes comprimidas como JPEGs, Acrobat Distiller las descomprimirá y volverá a comprimir (dependiendo de los parámetros que se hayan establecido). Esto dará lugar a la mencionada pérdida de calidad.

La extensión para Adobe Acrobat [Quite A Box Of Tricks](#) (de Quite Software) permite algoritmos de compresión JPEG más potentes que los del propio Acrobat. Además permite la recompresión de ficheros PDF.

## La compresión Zip

El algoritmo de compresión Zip es el que usan utilidades para PC muy populares como [WinZip](#), [PKZip](#), [Stuffit](#) y similares. su uso no quiere decir que Acrobat sea capaz de crear un fichero Zip, sino que usará el algoritmo para comprimir imágenes de color o escala de grises.

La compresión Zip es un poco más "inteligente" que la realizada por el algoritmo LZW. Lo que hace es escanear un fichero para detectar secuencias repetidas y entonces reemplaza éstas por una marca. Un ejemplo sería comprimir todo la *web* del Rincón del PDF y sustituir la palabra "fichero" (de gran frecuencia) por una simple marca. Una sucesión de sustituciones permite un gran ahorro de bytes.

Zip es un algoritmo de compresión "sin pérdidas". Esto quiere decir que las imágenes no se degradarán al comprimirlas. En Adobe Acrobat 4 es realmente así, pero en Acrobat 3 existe la opción de usar una compresión de 4 bits y otra de 8 bits. El problema es que con la de 4 bits, Acrobat reducirá primero el número de tonos por canal de 256 a sólo 16 y, posteriormente, realizará la compresión "sin pérdidas". El resultado es una tasa de compresión excelente y una degradación de la calidad extrema. Por consiguiente, la compresión Zip de 4 bits se debe evitar siempre a no ser que las características (color, contenido, etc...) de los ficheros se presten a ello. La compresión zip con 8 bits carece absolutamente de pérdidas.

## La compresión CCITT

La compresión CCITT se puede usar con imágenes de línea (en Photoshop: Mapa de bits). Es el mismo algoritmo de compresión que utilizan los faxes. Es una compresión "sin pérdidas", por lo que las imágenes no sufren degradación alguna.

Adobe Acrobat ofrece la posibilidad de elegir entre Grupo 3 de CCITT y Grupo 4 de CCITT. La mayoría de las opiniones están de acuerdo en que la segunda es mejor. Se puede dejar activada por omisión.



## La compresión RLE

El algoritmo RLE (Run Length Encoding) es un sistema de compresión "sin pérdidas", por lo que no afectará a la calidad de las imágenes. El algoritmo RLE es probablemente el más simple que existe. en vez de escribir un fichero que vaya diciendo: "Pixel blanco, pixel negro, pixel negro, ...." Va diciendo algo así como: "129 píxeles negros, dos píxeles blancos, ..."

En Acrobat, la compresión RLE se puede usar para imágenes de línea (en Photoshop: Mapa de bits). La mayoría de los usuarios prefieren la compresión CCITT porque es más eficaz.

## La compresión "hinchar/deshinchar" (flate)

El algoritmo de "hinchar/deshinchar" (*flate*, a veces también llamada *deflate*) es un sistema de compresión bastante complejo que he tratado con más profundidad en [otra página](#).

INICIO EL RINCÓN DEL PDF





# El texto en los ficheros PDF

[L. Leurs](#), 2000

(traducción: Gustavo Sánchez Muñoz)

El texto en un fichero PDF puede ir comprimido o sin comprimir. El algoritmo de compresión usado es [LZW](#). Si el texto se hubiera comprimido, ya no será legible a simple vista si se abre con un editor de texto.

El algoritmo LZW funciona reemplazando los datos más frecuentes por un sólo carácter. Suele dar una tasa de compresión de 2:1.

En los ficheros PDF el texto se distribuye línea a línea. Esto significa que un PDF no tiene "conciencia" de cómo fluía el texto en el documento original. Si usas Adobe Acrobat para hacer una pequeña corrección, las palabras no refluirán a la línea inferior si hiciera falta. Sin intentas seleccionar dos líneas en dos columnas adyacentes, verás que puedes seleccionar a lo largo de ambas columnas.

Afortunadamente, si copias y pegas las dos columnas en un documento de texto, verás que se pega en orden adecuado, una columna detrás de otra.

[Nota del traductor: Existen plug-ins para versiones de Acrobat anteriores a la versión 5 que permiten extraer el texto de los documentos. A partir de la versión 5, Acrobat mismo es capaz de guardar el texto de los documentos no bloqueados.]





# PDF frente a otros formatos de ficheros

[L. Leurs](#), 2000

(traducción: Gustavo Sánchez Muñoz)

Hay unos cuantos formatos de ficheros informáticos cuyos resultados se acercan a los de PDF. Éstos son algunos de ellos, con sus diferencias y similitudes:

## PDF frente a PostScript

Adobe, la compañía que desarrollo PostScript, fue también quien desarrolló PDF. De hecho, PDF está basado en lenguaje PostScript. PDF Usa el conjunto de intrucciones de PostScript de un modo diferente: Mientras que PostScript es realmente un lenguaje de programación que se puede usar para escribir un juego de ajedrez o un procesador de textos, PDF es de objetivos más limitados. Sólo describe la composición de un documento (usando operadores PostScript). En ese sentido, PDF recuerda más a una base de datos que a un lenguaje de programación.

Las principales ventajas de PDF frente a PostScript son:

- Los ficheros PDF tienden a ser más reducidos debido a los eficaces algoritmos de compresión que se pueden usar.
- Los ficheros PDF se pueden ver fácilmente con el lector Acrobat Reader u otras herramientas.
- Los ficheros PDF se pueden modificar más fácilmente, siempre que se tengan las herramientas apropiadas.
- Los ficheros PDF son má independientes de los dispositivos. Los ficheros PostScript suelen estar creados para dispositivos concretos y pueden generar errores PostScript si se envían a otros aparatos.
- Los ficheros PDF pueden ser más versatiler que los ficheros PostScript: Pueden contener enlaces a otros ficheros y elementos multimedia.

## PDF frente a HTML

PDF frente a HTML. El formato de datos...

PDF se suele comparar con HTML (el formato de datos más usual en la creación de páginas *web*). En principio HTML fue creado para describir la estructura de un documento y no tanto para elaborar su apariencia. El aspecto de una página *web* viene determinado por el navegador, no por el creador del documento. Con el incremento en la popularidad de la World Wide Web, el énfasis de las nuevas versiones de HTML pasó de más la estructura a la apariencia. En cierto modo, HTML se desplazó hacia los objetivos de PDF.

Al mismo tiempo, Adobe incluyó cada vez más funcionalidades *web* dentro de PDF (se hizo posible añadir enlaces de Internet dentro de un documento PDF). Al mismo tiempo, se puso a la disposición del público una extensión para los navegadores Netscape Navigator y Microsoft Internet Explorer para poder mostrar ficheros PDF. Acrobat 4 incorporó la capacidad de convertir un sitio *web* (o parte de él) en un documento PDF. Adobe añadió a PDF la capacidad de descarga incremental de bytes (*byte streaming*) para que no fuera necesario descargar todo un fichero PDF antes de poder ver su primera página en Internet.

En cierto modo, HTML y PDF se están convirtiendo en estándares en competencia. Ahora mismo, PDF es mucho más potente en lo que se refiere a describir la apariencia de documentos. HTML es más apropiado para accesos de Internet de baja capacidad. Sin embargo, es perfectamente factible usar ficheros PDF en sitios *web* y usar documentos HTML en un catálogo electrónico distribuido en CD-ROM.

## PDF frente a XML

XML (eXtensible Markup Language, "Lenguaje Extendido de Marcas"), es un formato que se puede usar para describir el contenido de un documento (similar a SGML, del que es un subconjunto). Ultimamente está en el centro de atención debido a que su gran flexibilidad permite una fácil integración con bases de datos y el intercambio de grandes masas de datos en Internet. En realidad, XML no compite con PDF. Mientras que XML describe el contenido de un documento, PDF describe su apariencia. No es fácil extraer el contenido de un fichero PDF (o al menos, no sin gran cantidad de trabajo manual) ya que la estructura de un documento se pierde cuando se transforma en fichero PDF.

Es muy interesante que PDF 1.3 incluya un mecanismo en forma de estructura en árbol que permite contener datos tipo XML. De este modo, es teóricamente posible que un fichero PDF contenga un resumen estructurado del contenido además de una representación exacta de su apariencia. Desgraciadamente, los

programas necesarios para incluir datos XML en un fichero PDF (usando marcas PDF: *pdfmarks*) como, por ejemplo, una Xtension para Quark XPress, aun no existen. Si necesitas PDF y XML, de momento el único camino es crear dos ficheros separados a partir de un programa de diseño o de un sistema de publicación de bases de datos.

Si quieres saber más sobre PDF y XML, hay un artículo muy interesante que Gordon Kent publicó en [PlanetPDF](#).

## PDF y Acrobat

Muchas personas confunden PDF (el formato de los datos) con Acrobat (el programa que la firma Adobe vende para generar, manipular y ver ficheros PDF). Esta confusión parece surgir del hecho de que casi todas las nuevas versiones de las especificaciones PDF van acompañadas de una nueva versión Acrobat. La versión 1.2 de las especificaciones PDF fueron acompañadas de Acrobat 3. Las especificaciones 1.3 de PDF, lo fueron de Acrobat 4.

INICIO EL RINCÓN DEL PDF





# Microsoft Word y el formato PDF

[L. Leurs](#), 2002

(traducción: Gustavo Sánchez Muñoz)

Muchas personas usan el formato PDF como una forma de imprimir profesionalmente documentos de Microsoft Office. Aunque esta es una idea bastante buena, hay que tener en cuenta algunas limitaciones de Microsoft Word.

La primera es que el uso de PDF-Writer es casi siempre una mala idea, al menos en lo que se refiere a impresión *profesional*. Lo recomendado es siempre imprimir a un fichero PostScript y después usar Acrobat Distiller o alguna herramienta similar para crear un PDF.

Si se instala Adobe Acrobat en un sistema que ya tenga instalada la *suite* Microsoft Office, se añaden macros a Word, Excel y Power Point que sirven para automatizar la creación de ficheros PDF. Siempre que se seleccione en estos programas la opción "exportar a PDF", la aplicación procederá a crear el fichero PostScript y después arrancará Distiller para convertir automáticamente ese código a PDF, lo que es algo bastante práctico.

## Problemas de Word con los PDFs

Estos son algunos de los problemas que existen. Buena parte de la información que aquí se explica fue amablemente proporcionada por Dov Isaacs ([Adobe](#)) y Jacob Schäffer ([Grafikhuset](#)).

1. Como la mayoría de las aplicaciones de Microsoft Office, Word se basa en el modelo gráfico del sistema sobre el que se esté ejecutando: GDI en Windows y QuickDraw en Macintosh. Al imprimir a un archivo PostScript, Word usa un controlador PostScript para generar el código correspondiente, salvo en el caso de los ficheros EPS, cuya información se limita a transmitir. Esto quiere decir que todos los colores del PDF resultante serán colores RGB y, aunque todos los RIPs PostScript más modernos pueden efectuar la separación de esos colores, el resultado podría no ser el esperado.
2. Ese código RGB resultante arriba mencionado hace los elementos negros o con un porcentaje de negro (grises) se filmen en las cuatro planchas (CMYK) de una cuatricromía. Para los Windows de la familia 9.x existe una versión (4.4.1 y posteriores) que incluye la opción de forzar la salida de los textos y gráficos de línea con el valor equivalente de Negro (dejando los colores restantes en sus valores RGB). Además, hay en el mercado varias herramientas y extensiones que permiten la conversión

de colores RGB a CMYK. Ese es el caso de [Quite A Box Of Tricks](#), (Quite, para PDFs ya hechos) y de [Colour Chameleon 2000](#) (Grafikhuset, para antes de hacer los PDFs).

3. Cuando trabaja y componen los textos, Word toma los valores de medida de las fuentes (*font metrics*) de las unidades de resolución del dispositivo de salida que tenga fijado en ese momento, y no de medidas absolutas de alta resolución. Eso quiere decir que cada vez que especifica una impresora nueva (aunque se siga usando el mismo controlador) el texto puede refluir y los saltos de línea pueden ser diferentes. Esto es aún más notable si se altera la resolución del dispositivo. Además, parece que los márgenes de impresión (es decir, la especificación de las zonas no imprimibles en el PPD y los ajustes del dispositivo) también tienen su influencia. Hay un par de medidas que se pueden tomar para evitar estos problemas:
  - Una solución es trabajar seleccionando el controlador de impresora de Acrobat Distiller y fijándolo a la resolución que se vaya a usar finalmente (1.200 ppp, por ejemplo). Para no alterar nada, las impresiones se deben hacer con los PDFs desde el menú de impresión de Acrobat.
  - En el menú "Herramientas - Opciones - Compatibilidad", Microsoft Word tiene una opción "Usar las medidas en la impresora para dar formato al documento" que hace que el texto tienda a refluir aún más. Esta opción se debe desactivar **antes** de crear cualquier documento.
  - En lo que se refiere al formateo de los textos, no es buena idea mover documentos de Word de una plataforma a otra (Windows a Macintosh o viceversa). Tampoco es buena idea llevarlos de Windows 9.x a Windows NT o Windows 2.000. En ambos casos puede haber redistribuciones, pero no es un arreglo fija.
  - Al fijar el interlineado de los textos es bueno evitar los interlineados automáticos ("sencillo", "multiple", etc...). Es mejor usar un valor fijo ("12 puntos", "14 puntos" ...).
4. Cuando se usan fuentes True Type con documentos de Word y se ha especificado en el controlador de impresora que las fuentes True Type deben enviarse a la impresora como "Fuentes de contorno", pudiera haber pequeñas diferencias en los espaciados de los caracteres tipográficos, especialmente en los finales de línea. Esto se puede solucionar diciéndole al controlador de impresora que incluya las fuentes True Type como fuentes de Tipo 42 (la persona que me informó de este problema, por cierto, estaba usando el PPD de Distiller 4 y el controlador PostScript Adobe PS 4.3.1)
5. A veces los ficheros de Microsoft Word incluyen gráficos creados en Microsoft Excel. Esta hoja de cálculo ofrece a los usuarios varios grosores de trazo de línea. Uno de ellos es "fino" (*hairline*), que lo que hace es decirle al dispositivo de impresión que debe trazar la línea con un punto de impresión de anchura (lo que no es 1 punto de pica, por cierto). En una impresora de 300 ppp de resolución, esto es una línea muy fina. En una filmadora de 2.400 ppp es una rayita invisible. Hay en el mercado varias herramientas y extensiones que permiten cambiar el grosor de las líneas. Una es la ya mencionada [Quite A Box Of Tricks](#), (Quite, para PDFs ya hechos). Otra es [PitStop](#) (de Enfocus Software).

6. En el menú "Herramientas - Opciones - Imprimir" existe una opción "Permitir alternar entre A4 y Carta" para redimensionar el documento en determinadas circunstancias. Esta opción se debe desactivar para que el documento no cambie al crear el PDF.

INICIO

EL RINCÓN DEL PDF

